Testing the Correlations between Beta Diversity and Productivity using WWF WildFinder and NASA MODIS Global NDVI Datasets

Jianting Zhang

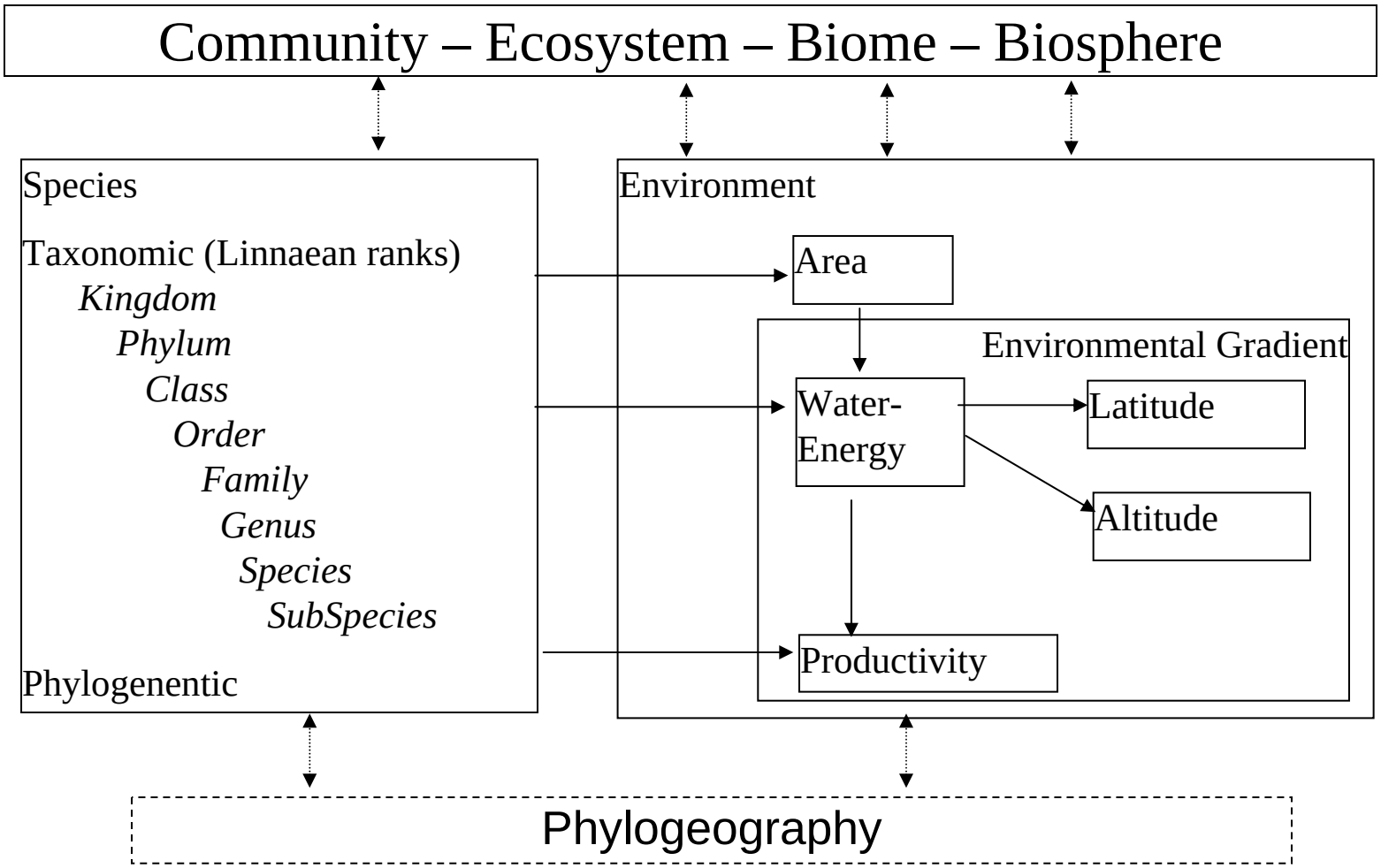
Dept of Comp. Sci. & NOAA-CREST

The City College of New York

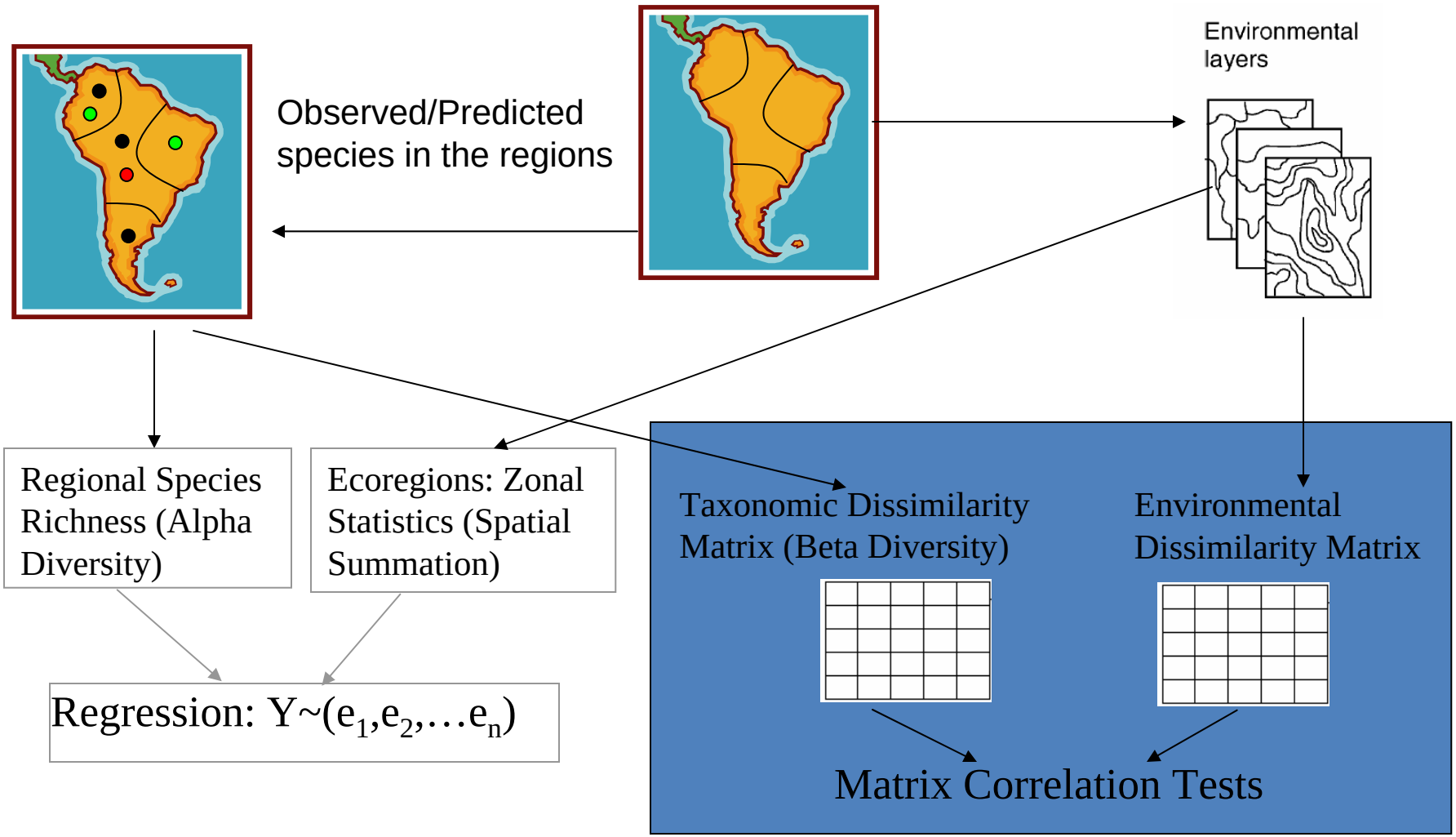
Introduction

- Exploring the relationship between species distribution and environments is central to basic ecological research and biodiversity conservation practices.
- Considerable amount of research on the relationships between species diversity and productivity at different spatial, ecological, and taxonomic scales has been conducted. However, the overall trend of the correlation at the global scale still remains sketchy.
- Satellite products provide spatially and temporally continuous coverage at large scales which is ideal for global scale studies. NDVI has been extensively used as a surrogate for primary productivity which motivates us to use MODIS NDVI data products for testing the correlation.

Introduction



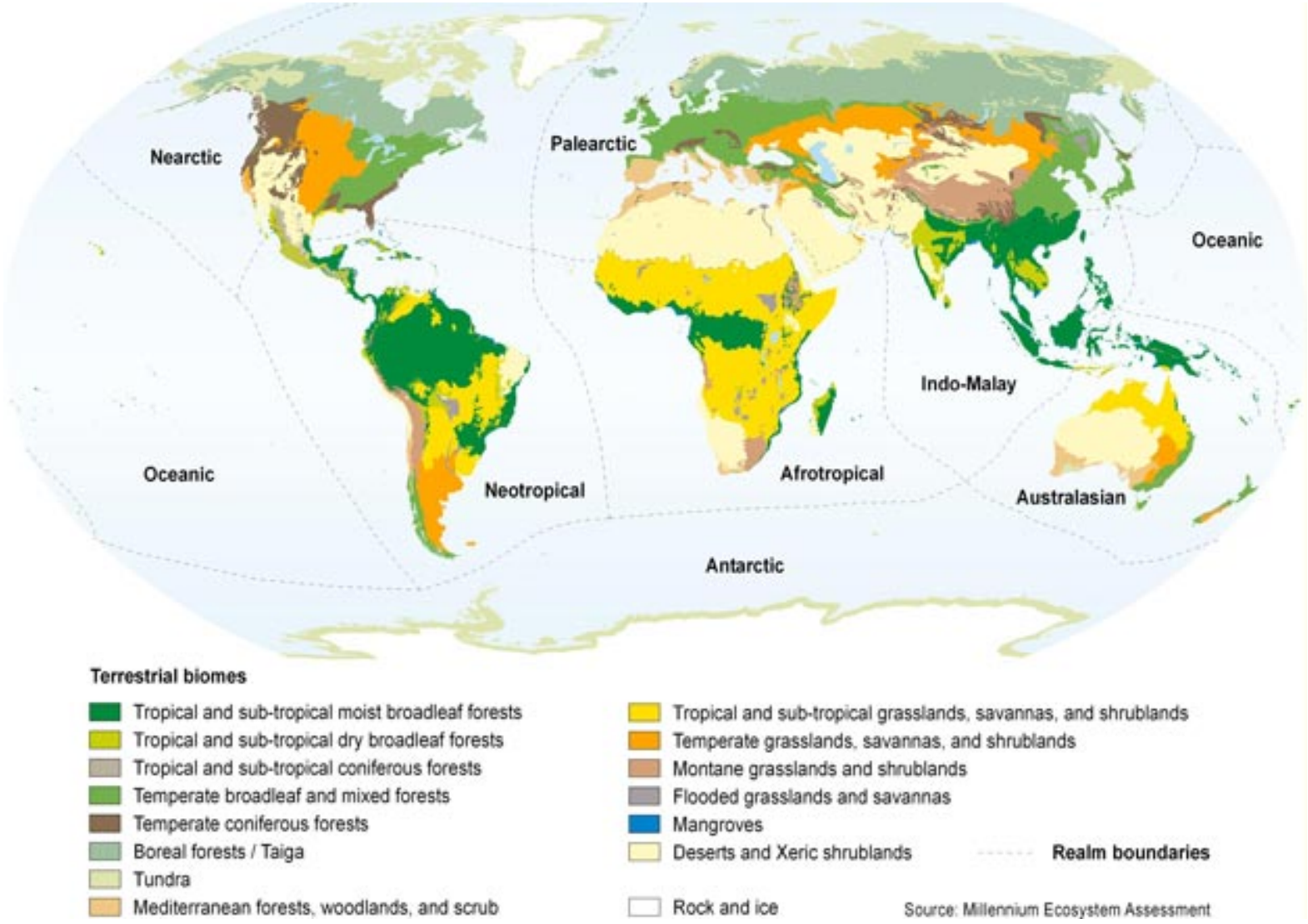
Introduction



Datasets

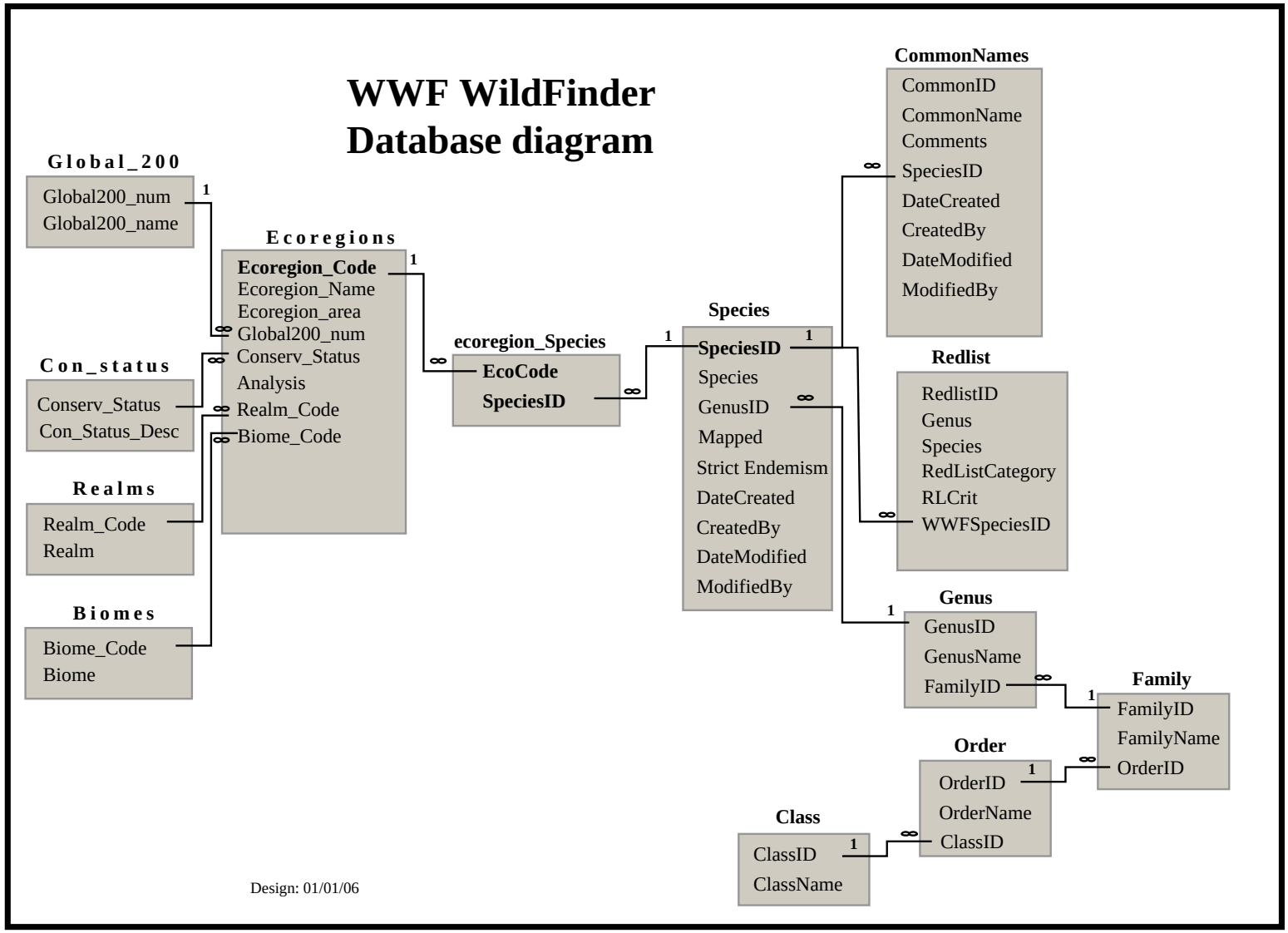
- WWF Terrestrial Ecoregions
  - 825 ecoregions
  - 8 biogeographical realms and 14 biomes
  - ~50M
- WWF WildFinder Database
  - 29,112 species, 4,815 genera, 445 families and 69 orders 4 classes (amphibians, reptiles, birds, and mammals)
  - 350,045 species-ecoregion records
- NASA NDVI
  - Filled Normalized Difference Vegetative Index (NDVI)
  - Global data set of spatially complete NDVI maps for 23 sixteen-day periods per year
  - 2000-2004 average
  - 1 Minute (~ 2km) resolution (21600*10800)

Global Biomes and Realms



Source: http://www.worldwildlife.org/science/data/item1875.html (Olson et al 2001)

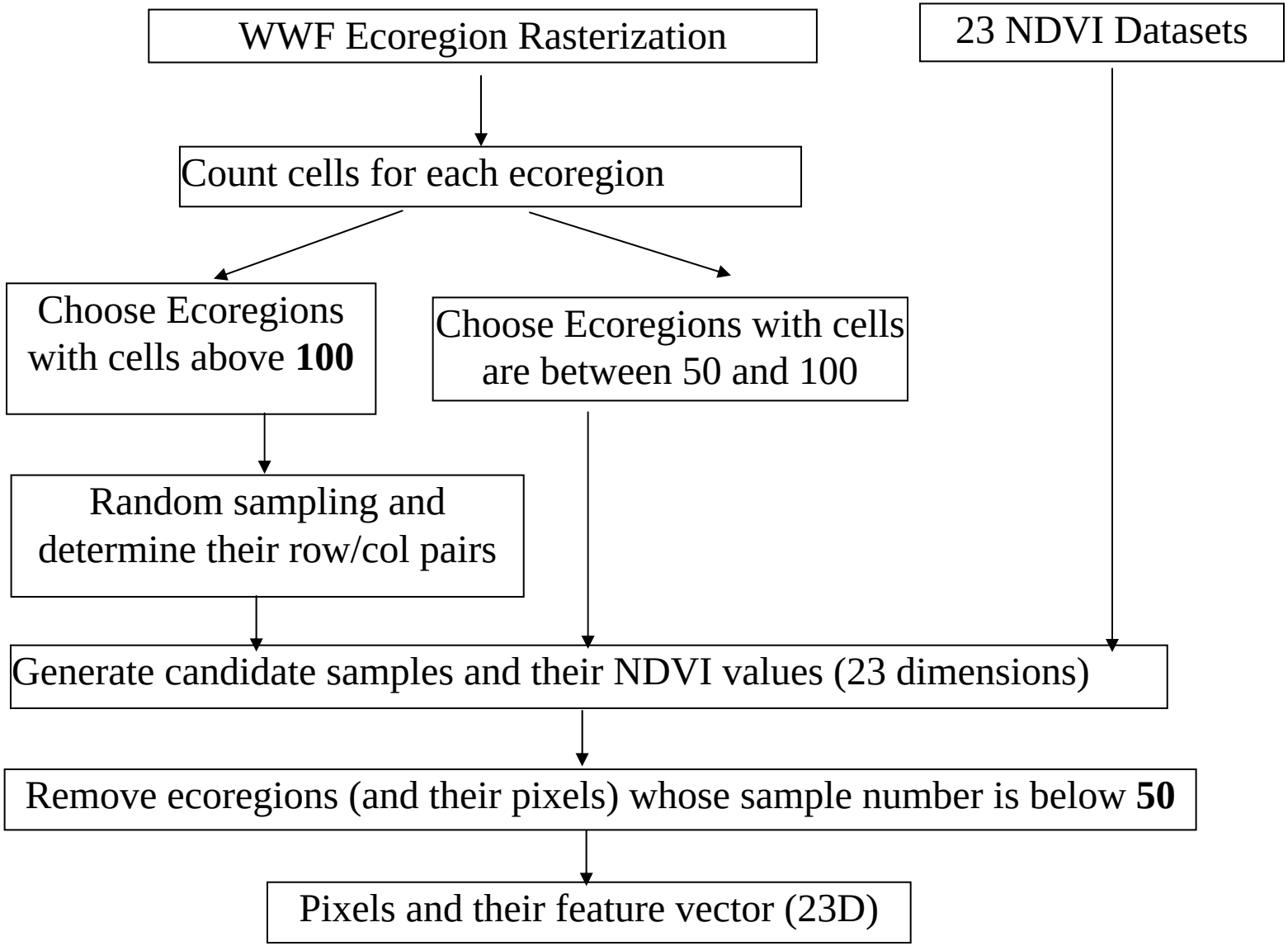
WWF WildFinder Database



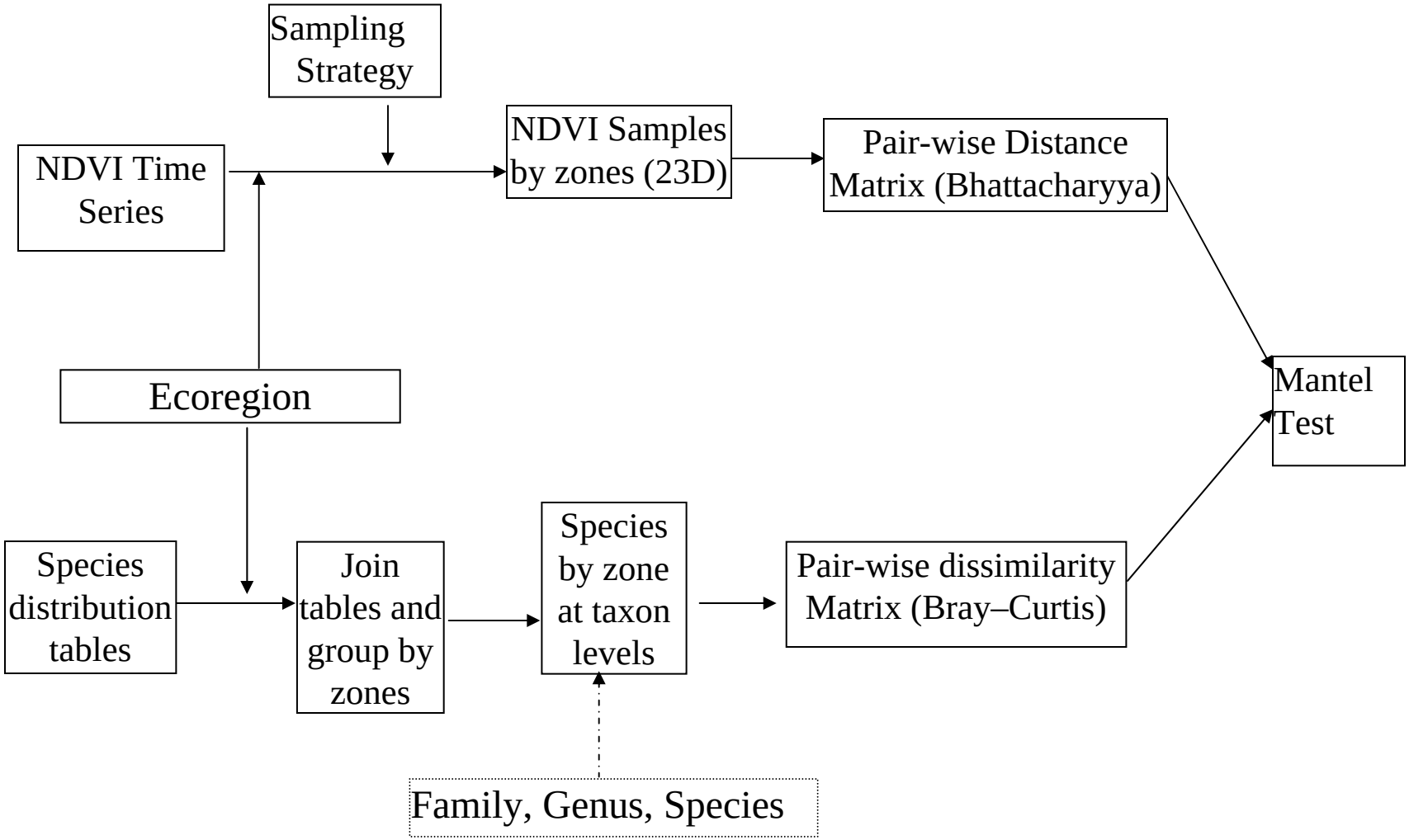
Preprocessing of NDVI Dataset

- Data format conversion: HDF4->GeoTIFF
  - TIFF is better supported
  - Build GDAL driver for HDF4
- Retrieval of time series NDVI values for sampled pixels
  - GDAL RasterIO (C++ API)
  - Sampling strategy will be provided shortly

Pixel Sampling Strategy



Method


Sampling Strategy
NDVI Time Series
NDVI Samples by zones (23D)
Pair-wise Distance Matrix (Bhattacharyya)
Ecoregion
Mantel Test
Species distribution tables
Join tables and group by zones
Species by zone at taxon levels
Pair-wise dissimilarity Matrix (Bray–Curtis)
Family, Genus, Species

Method

- Distances

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \qquad \Sigma_{ij} = \mathrm{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big]$$

$$\begin{matrix} x_{11} & x_{12} & x_{13} & .. & x_{1n} \\ x_{21} & x_{22} & x_{23} & ... & x_{2n} \\ & & & & \\ & & & & \\ x_{m1} & x_{m2} & x_{m3} & & x_{mn} \end{matrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_1 - \mu_1)(X_n - \mu_n)] \\ \mathrm{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \mathrm{E}[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mu_n)(X_1 - \mu_1)] & \mathrm{E}[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & \mathrm{E}[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}.$$

Euclidean Distance $dE = \sqrt{(\mathrm{x_i} - \mathrm{x_j})^{\mathrm{T}} (\mathrm{x_i} - \mathrm{x_j})} \quad \sqrt{\sum_{\mathrm{k=1}}^{\mathrm{n}} (\mathrm{x_{ik}} - \mathrm{X_{jk}})^2}$

Mahalanobis Distance $dM = \sqrt{(\mathrm{x_i} - \mathrm{x_j})^{\mathrm{T}} \sum^{-1} (\mathrm{x_i} - \mathrm{x_j})}$

Farber, O. and R. Kadmon (2003). "Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance." Ecological Modelling 160(1-2): 115-130.

Method

Bhattacharyya Distance (Bhatt):

- Measures dissimilarity between two groups of samples instead of two samples
- Takes both mean and variance into consideration

$$dB = \frac{1}{8}(\mu_i - \mu_j)\left(\frac{\sum_i + \sum_j}{2}\right)^{-1}(\mu_i - \mu_j) + \frac{1}{2}\ln\left(\frac{\left|\frac{\sum_i + \sum_j}{2}\right|}{|\sum_i|^{1/2}|\sum_j|^{1/2}}\right)$$

Bray–Curtis Dissimilarity $\beta_{BC} = \frac{\sum_{i=1}^{n} | x_{ai} - x_{bi} |}{\sum_{i=1}^{n}(x_{ai} + x_{bi})}$

For presence/absence data:(b+c)/(2*a+b+c), complementary measurement of the Sørensen similarity

Koleff, P. et al. 2003. Measuring beta diversity for presence-absence data. - Journal of Animal Ecology 72: 367-382.

Method

Mantel Test of Matrix Correlation

$$z = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{1\,ij} d_{2\,ij} \qquad r = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(d_{1\,ij} - \overline{d_1})}{S_1} \frac{(d_{2ij} - \overline{d_2})}{S_2}$$

Permutation Test

- Permute all independent pairs (possibly through sampling) 20!=2432902008176640000
- Using the rank of the test score among the scores of all independent pairs as the significance level
- If a z/r value ranks #50 among 10000 tests, then the confidence level is p=50/10000=0.005

- Vegan package in R

Results –Global

	Cor.	Sig.
NDVI-Family	0.06797	<0.00001
NDVI-Genus	0.12839	0.0001
NDVI-Species	0.12872	<0.0001

Results –Global Realms

Realm	No. of Ecoregions	NDVI-Family		NDVI-Genus		NDVI-Species	
		Cor.	Sig.	Cor.	Sig.	Cor.	Sig.
AA	80	0.02119	0.3294	0.24165	<0.0001	0.26665	<0.0001
AT	105	0.11857	0.0132	0.20153	<0.0001	0.23469	<0.0001
IM	102	0.20775	0.0030	0.36345	<0.0001	0.31556	<0.0001
NA	108	0.33208	<0.0001	0.33479	<0.0001	0.38295	<0.0001
NT	168	0.20295	0.0001	0.24989	<0.0001	0.23283	<0.0001
OC	12	0.45397	0.0025	0.41631	0.0039	0.44092	0.0008
PA	187	0.27670	<0.0001	0.32669	<0.0001	0.28259	<0.0001

Results –Global Biomes

Biome	No. of Ecoregions	NDVI-Family		NDVI-Genus		NDVI-Species	
		Cor.	Sig.	Cor.	Sig.	Cor.	Sig.
01	212	0.01559	0.2779	0.09360	<0.0001	0.13017	<0.0001
02	53	0.08033	0.0840	0.16257	0.0007	0.19261	0.0001
03	16	0.26428	0.0697	0.27841	0.0523	0.27937	0.0551
04	82	0.16130	0.0097	0.18296	0.0004	0.19489	<0.0001
05	53	0.25642	0.0001	0.25675	<0.0001	0.28333	<0.0001
06	28	0.23048	0.0237	0.22027	0.0074	0.19028	0.0167
07	43	-0.03457	0.6420	0.01021	0.4247	0.08205	0.1133
08	41	0.31976	<0.0001	0.32009	<0.0001	0.29928	<0.0001
09	21	0.33059	0.0001	0.32888	<0.0001	0.33468	<0.0001
10	50	0.15550	0.0088	0.22306	<0.0001	0.23951	<0.0001
11	14	0.72931	<0.0001	0.62436	0.0006	0.51430	0.0024
12	39	0.23208	0.0004	0.22302	0.0007	0.23097	<0.0001
13	92	0.33580	<0.0001	0.32663	<0.0001	0.30910	<0.0001
14	19	0.00329	0.4444	0.03544	0.2691	0.03003	0.3266

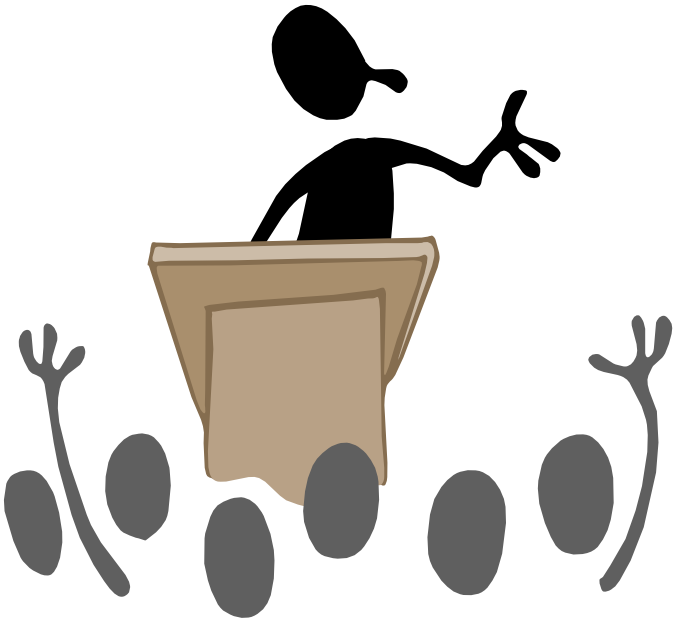
Summary

- This presentation focuses on:
  - Data processing: practical skills to integrate multi-source datasets and handle large-volume datasets
  - Reporting preliminary results
- Further work to interpret the results and validate the methodology is needed:
  - Work closely with ecologists, biogeographers and statisticians.
  - Identifying regions that have high positive or high negative correlations – automatic detection.

Ongoing Works

- Rasterizing Species Range Maps from NatureServe
  - Mammal (1693), Birds (4148), Amphibian (5816)
  - 1 minute spatial resolution, (360*60)*(180*60)=23,328,000 cells for each of the 11,657 species
- Designing new tree data structures for efficient representations + storing the equivalent linear data structures in PostgreSQL=> efficient query processing at the sub-polygon level
- Planned ecological/biogeographical research
  - Dissimilarity matrix correlations at the finer and multiple scales
  - Spatial correlations for both environmental and species data
  - How different are WWF Wildfinder and NatureServe data?
  - Are current ecoregion systems function as expected?
  - ……

Q&A



jzhang@cs.ccny.cuny.edu